Tools for text mining over biomedical literature

Rinaldi, Fabio ; Schneider, G ; Kaljurand, K ; Hess, M

Abstract: This poster describes OntoGene: an environment, based on a deep linguistic parser, aimed at supporting the process of Text Mining from Biomedical Scientific Literature. We will illustrate in particular the Relation Extraction component and the development support facilities.

An Environment for Relation Mining over Richly Annotated Corpora: the case of GENIA

Fabio Rinaldi[*], Gerold Schneider[*], Kaarel Kaljurand[*], Michael Hess[*], Martin Romacker[**]

Abstract

The biomedical domain is witnessing a rapid growth of the amount of published scientific results, which makes it increasingly difficult to filter the core information. There is a real need for support tools that 'digest' the published results and extract the most important information.

We describe and evaluate an environment supporting the extraction of domain-specific relations, such as protein-protein interactions, from a richly-annotated corpus. We use full, deep-linguistic parsing and manually created, versatile patterns, expressing a large set of syntactic alternations, plus semantic ontology information.

[*]Institute of Computational Linguistics, IFI, University of Zurich, Switzerland, {rinaldi, gschneid, kalju, hess}@ifi.unizh.ch

[**]Novartis Pharma AG, Basel, Switzerland martin.romacker@novartis.com